Study ID	Description of cohort including numbers	confounding variables collected and notes on matching/ adjustments made	Alternative exposure group selection feasible	Comments
Cohort studies				
Fleming 1978	RCGP Outcomes of Pregnancy study 1975: 9,000 women; from this was selected a random sample of 500 pregnancies proceeding to normal outcomes	No confounding variables reported.	Yes†	Doubtful malformation (n =140) and rhesus incompatibility (n=37) group and stillbirth (n =100) not used
Goujard 1979	3,379 women recruited in Obs & Gynae clinics in Paris and Lille, France during pregnancy. from 1975 to 1977	Information on their current pregnancies including symptoms and the medicaments taken; previous pregnancies and their general health backgrounds.	No	
Hadjigeorgiou 1982	Maternity Hospital Greece, births 1975-77, 15,535 live births, and 559 exposed to sex hormones), 14,976 no hormones,	Cytomegalovirus, infection, toxoplasmosis, hepatitis, syphilis, rubella, teratogenic drugs. Diseases and medication reported at admission prior to birth.	No	
Haller 1974	3588 pregnant women, recruited Oct 1969 to April 1972, University Hospital Göttingen; 617 with hormonal pregnancy test	No confounding variables reported High number of induced abortions 13/154 (8.4%)	No	
Kullander & Kallen 1976	6,376 pregnancies, Malmo, 1963-5, resulting in 5,753 live births, 5,002/5,753 non malformed, 751/5,753 malformed. 156 women took Primodos.	Age, parity, Civil status, social class and drugs were recorded. The group of induced abortions containing unwanted pregnancies, shows a high incidence of Primodos use (X^2= 12.9 P<0.001).	No	
Meire & Vuylsteek 1978	500 mothers consecutive births in 3 hospitals in Bruges, Belgium, 20 had taken HPTs.	No confounding variables reported.	No	
Michaelis 1983	13,643 pregnancies	Detailed general and gynaecological history, drug intake, exposure to chemical agents, daily workload, intercurrent diseases, accidents, surgical operations and other factors. Unmarried women and women who became married during the present pregnancy were much more frequent amongst the Duogynon takers than in the total study. Controls matched according to the marital status. Smoking habits and frequency of medications used during the pregnancy were similar in Duogynon takers and their controls. For all subgroups, there were no major differences in the frequencies of reported previous diseases.	Yes†	We selected Duogynon without other hormones (n=502) and excluded Duogynon with other hormones (n =108)
Roussel 1968 unpublished	Data on 1582 pregnancies 1966 to 1967 were complete	No confounding variables reported. No difference in incidence of mothers over 30 compared with under 30 for number of CNS malformations (X2 0.644); no difference in age of groups; no evidence for increase found in 2nd baby following birth of a baby with CNS malformation (X^2 = 0.0209).	No	
Rumeau-Rouquette 1978	1963-69, recruitment in 12 gynaecology clinics in Paris; 12,764 women gave birth to 12,895 children in hospitals participating in study; controls were mothers of unaffected infants selected at random in the same hospital	Medical history, course of pregnancy, infectious diseases, inoculations, reproductive history, social and occupational category, use of alcohol, tobacco Of the initial cohort recruited, 5,511/18,275 dropped out during first 9 weeks pregnancy (presumably a high proportion due to miscarriage but this is not stated in the paper)	Yes†	other use of oestrogen progestogen derivatives that were not hormone pregnancy tests (n=1224)
Torfs 1981	19,906 full term pregnancies, 227 of which exposed to HPTs.	Age, medical and reproductive history, socio economic information, ethnicity. The final analysis removed from our comparison groups all pregnancies for which an oestrogen/progestogen was given	Yes†	we did not select the serum group (n=689) or the urine test (n=332) group as controls: differences in

		for reasons other than a pregnancy test; these reasons include threatened abortion as well as other conditions.		SCA and NSCA between hormone test and either serum or urine tests were not significant
Case control studies				
Ferencz 1980	Mothers of 110 infants with conotruncal malformations of the heart, born 1972-75.	Maternal health (hospitalisations, illnesses, therapies,); past reproductive history; index pregnancy factors including contraception used previously, fertility treatments, symptoms during pregnancy, illnesses during pregnancy, medications during pregnancy including hormones; smoking; alcohol intake; occupational history of mother and father; exposure of mother to fumes, paints and insecticides; family history including history of congenital malformations in previous children or in close relatives. For each case, three normal controls were chosen from the birth population: 2 matched on eight characteristics related to the likelihood of hormone-taking (race, maternal age, parity, fetal losses, gestational age, delivery mode, time of prenatal registration, private service), and one also on the infant's sex and birthweight; the third control was chosen at random. We extracted data on the most matched controls.	Yes†	Matched controls (n=186) and random controls (n=110); (n= 20) disease controls not included - non used hormone pregnancy test. For characteristics presumed to be predisposing to congenital malformations, risk scores were constructed to be used as interaction variables. No significant differences were noted between groups. In the analysis of cases versus matched controls no effect modification was seen by any risk variable, this analysis suggests that the relative risk is unity uniformly throughout the population of cases.
Gal 1972*	100 mothers of infants with spina bifida, and controls	Age, parity, reproductive history, illnesses, illegitimacy, bleeding. Matched for week of baby's birth; age of mother (5 year bands), reproductive history, course of pregnancy, sex of baby.	No	
Greenberg 1977**		Antenatal, personal, and family history and drugs prescribed during the first trimester. Controls: babies born within 3 months of and based at the same general practice as matched cases.	Yes†	Cases and controls matched for all factors except history of previous children with abnormalities in the study families
Hellstrom 1976	32 patients with congenital limb abnormalities born during 1965-74 and 30 controls with spina bifida born during the same period.	Maternal age, parity. Controls were children with spina bifida, therefore scored selection of controls as inadequate. Other of the cases had been exposed to other exogenous hormones during pregnancy (4 "treated with hormone").	No	
Lammer & Cordero 1986	1,091 mothers of infants with abnormalities born 1 July 1970 to 20 June 1979, (21% not completed data collection)	Race, maternal education, family history, socio-economic status, parity, previous fetal loss. Control group was composed of infants with malformations other than the one under investigation. e.g. for spina bifida, controls were those with non-spina bifida abnormalities.	No	Choice of controls as infants with other major malformations could underestimate the true association.
Laurence 1971	1968-1970, UK;	No confounding variables reported. "in London the controls were the next baby with no abnormality born in the same hospital; in Exeter, control mothers were matched for area of birth, parity and month of conception; in Wales the control mothers were those who had had one baby with spina bifida or anencephaly and had a subsequent normal birth during the study period and as such these were not matched invidually."	No	
Levy 1973	76 cases, 76 controls	No confounding variables reported, Controls were infants with Mendelian disorders, matched for date of birth.	No	
Nora 1978 Case control 1	32 patients with VACTERL, 60 controls	Age, date of birth, sex, gestational age, race, socioeconomic levels, areas of residences, parity. Matched as closely as possible for age, date of birth, sex, gestational age, race, socioeconomic levels, areas of residences, parity. no mendelizing disorders were included	No	Probable underestimate because HPT use is clear for cases but not for controls therefore we took the

				upper limit for controls, which includes all hormone use, (n=5)
Nora 1978 Case control 2 and 3	236 patients with full variety of cardiac lesions, 412 controls with known single mutant gene and chromosomal disorders	Sex, race, approximate date of birth, area of residence. Matching was for sex, race, approximate date of birth, area of residence. Reported controls with birth defects were chosen because mothers recall the histories better than those of their normal off spring	No	eliminating cases from both groups in which there were known potential teratogens in addition to progestogens and oestrogens also led to a highly significant difference p<0.001)
Polednak 1983	99 singleton male births with hypospadias and 99 matched controls	Parity, maternal age, race, area of residence Most adjacent birth date, matched for maternal age, race, area of residence	No	
Rothman 1979	390 cases, 1254 controls. HPTS: 14/388 cases vs 35/1246 controls	Parity, mother's education level, insulin use, alcohol, tobacco. Births within same 3 years of the study period	No	Strongest association among the other drugs reported was for insulin
Sainz 1987	Cases identified via the national collaboration of 42 hospitals registering congenital abnormalities between April 1976 to Sept 1984, and controls matched for sex, hospital. 244 cases. 8735 non-affected birth following birth of a malformed child of same sex and born in same hospital. Matched by sex, date of birth, hospital of birth	Matched by sex, date of birth, hospital of birth. Same sex, matched for data of birth, born in the same hospital, no abnormality. In our study, the cases and controls come from the same hospital and they have been compiled during the same period. The percentage of exposed and non-exposed children in each group does not depend on medical preferences as the same obstetricians have treated the mothers of the cases and controls.	No	
Nora 1975	15 patients with multiple congenital anomalies. 30 controls (15, with chromosomal anomalies, q5 with functional murmurs	Matched for age, 15 with chromosomal abnormalities, 15 with functional murmurs. 2 cases with exposure to potential teratogens were eliminated	No	likely underestimate as HPT use is clear for cases but not for controls therefore we took the upper limit for controls, including all hormone use: n=3
Janerich 1977	104 cases with birth certificate mentioning CHD, 104 matched controls	Age, county of residence, dob, race, medications, infections. From adjacent birth record matched by mother's age, county of residence, dob, race	No	No evidence that the use of other medications during the first trimester weakened the strong association
Tummler 2014	296 cases, 3,676 malformed infants	No confounding variables reported. Data from the Malformation Monitoring Centre Saxony-Anhalt, Germany. 3676 malformed infants served as the control group	No	
Janerich 1974	108 cases with congenital limb defects and 108 healthy controls	Controls matched on birth date, mother's race and age +/- 2 years; and by default due to adjacent records for cases and controls these matched well on county of residence of the mothers.	No	cases had a higher number of births, and a slight increase in twins

* Excluding cases of previous malformed babies and history of infertility did not affect the significance: cases 15/85 vs control 4/97 (p=0.01 to 0.001).
Differences were found with regard to maternal age and acute infection

** Additional analysis: excluded all case-control pairs who had a family history of congenital malformation in either or both families; excess use of HPT by case mothers remains statistically significant: cases 64/743 vs control 35/781 (X^2 9.42; $P < 0.01$) cases and controls matched for all factors except history of previous children with abnormalities in the study families

† Of the 16 cases controls studies we did not included data from 2 studies in the analysis where there was the potential to select an alternative group for comparison. We therefore did not use 0.23% of potential available control data (40/17095) for the following reasons.

Ferencz 1980 :

- Matched controls (n=186) and random controls (n=110) included in the analysis;
- Disease controls (n =20) not included in the analysis (non used hormone pregnancy test).

Greenberg 1977:

- Matched controls (n=836) included in the analysis
- Data from (n =20) stated as exposed to hormones as both case and controls (likely these were exposed twins or family members) were not included in the analysis.

† Of the 10 cohort studies we did not included data from 4 studies in the analysis where there was the potential to select an alternative group for comparison. We therefore did not use 5.61% of the available data (3132/55974) for the following reasons.

Fleming 1978:

- Total affected with lethal and unequivocal malformation (n=245), total non-affected (controls n=500) included in the analysis.
- Doubtful malformation (n =140) and rhesus incompatibility (n=37) group and stillbirth (n =100) not used in the analysis

Michaelis 1983:

- Duogynon without other hormones (cases n=502); total non-affected (controls n=502) included in the analysis.
- Duogynon with other hormones (n =108) not included in the analysis.

Rumeau-Rouquette 1978:

- Total affected (cases n= 1,150); total non-affected (controls n = 9,822) included in the analysis.
- Other use of oestrogen progestogen derivatives that were not hormone pregnancy tests (n=1224) not included in the analysis.

Torfs 1981:

- Total affected (cases n=203); total non-affected (controls n =17,057) included in the analysis.
- Serum group (n=689) or the urine test (n=332) group not included in the analysis: